Subjectivity in chromosome band–level estimation: a multicenter study

Katherine B. Geiersbach, MD[1,2], Anna E. Gardiner, BS[3], Andrew Wilson, MS[3],
Shashirekha Shetty, PhD[4], Hélène Bruyère, MD[5], James Zabawski, MS[6], Debra F. Saxe, PhD[7],
Rebecca Gaulin, BS[8], Cynthia Williamson, BS[9] and Daniel L. Van Dyke, PhD[9]

Purpose: Chromosome band level is the primary quality indicator for G-banded metaphase chromosome analysis. Although current professional guidelines address the minimum necessary band level for constitutional studies, there is no study documenting the comparative performance of different band-level estimation methods.

Methods: This study compared 5 band-level estimation methods (Stallard, Vancouver, Welborn, United Kingdom External Quality Assurance Scheme, and Ford) in a multicenter study in which 82 readers from 7 different clinical cytogenetics laboratories evaluated the same 10 karyotypes (5 from amniotic fluid and 5 from peripheral blood) by each method.

Results: There was a 94% correlation between the five band-level estimation methods. The Welborn method yielded significantly lower

scores for amniotic fluid karyotypes ($P < 0.01$) but not for peripheral blood karyotypes ($P = 0.75$). The distribution of scores obtained from different readers suggests a high level of subjectivity in chromosome band–level assessment. The variation in band-level estimation did not correlate with reader experience or study center, except for readers from one laboratory, for which the distribution of scores was significantly lower ($P < 0.01$).

Conclusion: The results from this study suggest that the consistent use of one method is more important than the actual method employed for monitoring karyotype quality.

Genet Med advance online publication 25 July 2013

Key Words: band level; band resolution; chromosome analysis; cytogenetics; karyotype

INTRODUCTION

Chromosome analysis using G-banding is the oldest and most widely used method for whole-genome analysis. It is also the most comprehensive method for analyzing the large-scale architecture of the entire genome, including highly repetitive sequences comprising the pericentromeric regions, telomeres, and acrocentric short arms—regions that are not well covered by newer technologies such as chromosomal microarray or next-generation sequencing. To ensure the most detailed chromosome analysis, cells in early metaphase are preferable because of their higher band level. Laboratory methods designed to enhance the selection of early metaphase cells include synchronization of cell cultures and the use of intercalating agents such as ethidium bromide.[1] For the detection of chromosomal abnormalities, the best-quality metaphase cells are selected to allow for the finest discrimination of break points and the detection of smaller aberrations. Estimation of band level is typically used to assess the quality of a G-banded metaphase cell. With an increasing number of bands per haploid set (bphs), a more detailed evaluation can be performed,

and smaller aberrations can be detected. The American College of Medical Genetics and Genomics recommends that "the 550-band stage should be the goal of all peripheral blood studies," and "a minimum of 400 bands should be reached for 90% of analyses from amniotic fluid and chorionic villi cells."[2] The College of American Pathologists Laboratory Accreditation Program requires a 400 band level for constitutional cases and at least a 550 band level "in appropriate blood samples" (such as those evaluated for intellectual disability, dysmorphology, and birth defects).[3] The latter recommendation (a 550 band level for "appropriate blood samples") has been superseded by chromosomal microarray as the recommended first-tier test for this population.[4,5] For evaluation of acquired chromosome abnormalities, chromosome analysis remains the standard of care, although neither the American College of Medical Genetics and Genomics nor the College of American Pathologists has recommended a minimum band level. For both constitutional and acquired abnormalities, G-banded chromosome analysis remains widely practiced, and the ultimate measure of the quality of chromosome analysis is the band level. The clarity of

staining, straightness of the chromosomes, and lack of overlaps with other chromosomes in a metaphase spread also affect the quality of chromosome analysis, but these factors are seldom recorded during standard chromosome analysis.

Idiograms for band levels of 300, 400, 550, 700, and 850 bphs are published in *ISCN 2013: An International System for Human Cytogenetic Nomenclature*.[6] On average, a 400 band level allows discrimination of copy-number changes of ~9 Mb or higher, and a 550 band level allows discrimination of copy-number changes of ~6 Mb or higher.[7] However, the ability to detect smaller aberrations by cytogenetic analysis is influenced by additional factors, such as whether the aberration alters the banding pattern and the density and pattern of bands in the region of interest.

Comprehensive band-level assessment by exhaustive counting of all the chromosome bands is not practical for routine clinical studies. A number of alternative band-level assessment methods exist. One of the oldest methods is the Stallard method, which employs a count of bands on chromosome 10 to determine the overall band count of a cell.[8] The "Vancouver method," published in 1991, expands on this method by including a count of bands on chromosome 10 as well as four other chromosome regions (1p31–32, 11p, 12q, and X) to achieve a more accurate band count.[9] In 1993, Jeanna and Roger Welborn[10] published a comparison of several methods for estimating the true band count; the most accurate and simple of these methods was the total count of bands on chromosomes 1 and 2 multiplied by a factor of 6, because these two chromosomes comprise approximately one-sixth of the human genome. An alternative to band counting is the evaluation of the presence or absence of target bands to estimate band level. Two well-described target methods exist. The first was developed by the Association for Clinical Cytogenetics and is utilized by the United Kingdom External Quality Assurance Scheme (UKEQAS).[11] The second was developed by Zabawski et al.[12] at Henry Ford Hospital and published in 2005. The UKEQAS method is a standard approach used by laboratories in much of Europe, but there is no standard method currently in use in the United States. In addition, we are unaware of any published side-by-side comparison of all of these methods to assess their relative performance. In the current study, we have compared the five methods (Stallard, Vancouver, Welborn, UKEQAS, and Ford) in a multicenter study in which 80 readers evaluated the same 10 karyotypes (5 from amniotic fluid and 5 from peripheral blood) by each method.

MATERIALS AND METHODS
Karyotype preparation
Cases were processed in a Clinical Laboratory Improvement Amendments–certified clinical laboratory environment (ARUP Cytogenetics Laboratory, Salt Lake City, UT) according to standard cytogenetic laboratory techniques. Briefly, peripheral blood samples were grown in suspension cultures for 48–72 h using a lymphocyte mitogen (phytohemagglutinin), a spindle apparatus inhibitor (colcemid), nucleoside

analogs (amethopterin and thymidine), and an intercalating agent (ethidium bromide) as previously described.[1] Amniotic fluid samples were grown in monolayer cultures on glass coverslips and were treated with colcemid before harvesting. After culture and harvest, G-banding was performed using trypsin and Wright staining. Slide preparations of peripheral blood and slide mounted coverslips from amniotic fluid cultures were digitally scanned, and metaphases were karyotyped using Metafer and Ikaros software (MetaSystems, Altlussheim, Germany).

Karyotype selection and distribution
Digitally archived peripheral blood and amniotic fluid cases were reviewed sequentially to select karyotypes in which all regions designated for evaluation were free from overlaps or other artifacts on both chromosome homologs. Five peripheral blood karyotypes and five amniotic fluid karyotypes were used for the study. The selected karyotypes were de-identified, and printed hard copies were distributed to each study center, along with a survey and detailed scoring instructions. Readers were asked to designate their experience in clinical cytogenetics in years and months and to complete a written score sheet for each of the 10 karyotypes. A total of 82 readers from seven cytogenetics laboratories (ARUP Cytogenetics Laboratory, Mayo Clinic Cytogenetics Laboratory, Henry Ford Health System Cytogenetics Laboratory, Cleveland Clinic Cytogenetics Laboratory, Vancouver General Hospital Cytogenetics Laboratory, Emory Clinical Cytogenetics Laboratory, and University of Washington Cytogenetics Laboratory) participated in the study. Four readers participated in only half of the study (two readers completed only the amniotic fluid section and two readers completed only the peripheral blood section). The total number of readers for each karyotype was 80 readers. A total of 4,000 scores were generated (80 readers scored 5 amniotic fluid karyotypes by 5 different methods and 80 readers scored 5 peripheral blood karyotypes by 5 different methods). The study protocol was approved by the institutional review board of the University of Utah.

Band-level assessment methods
Readers were asked to designate a separate score for each chromosome homolog evaluated, and the lower of the two scores was used to compute the band level, as detailed below. Cutoff thresholds were 400 bphs for amniotic fluid karyotypes and 550 bphs for peripheral blood karyotypes.

Stallard method
The total number of dark and light bands on chromosome 10 was used to compute the band level. According to this method, a count of 13 or more bands indicates a band level of at least 400 bphs, and a count of 26 or more bands indicates a band level of at least 550 bphs.[8]

Vancouver method
This method was a sum of the total number of dark bands on multiple regions: 1p31–p32, 10, 11p, 12q, and X (see

Supplementary Figure S1 online). Band counts for chromosome 10 were derived from the results from the Stallard method (total count divided by 2, rounded down to nearest whole number) in an effort to reduce the overall time investment required from survey participants. According to the published criteria for this method, a total score of 21 or more dark bands for all regions (1p31–p32, 10, 11p, 12q, and X) indicates at least 400 bphs, and a score of 40 or more dark bands indicates at least 550 bphs.[9]

Welborn method

The total number of dark and light bands on chromosomes 1 and 2 was multiplied by a factor of 6 to compute the band level.[10] This method provides an absolute band count, and therefore no conversion is needed; hence, the absolute thresholds of 400 bphs for amniotic fluid and 550 bphs for peripheral blood were used.

UKEQAS method

Readers were asked to indicate whether target bands were present or absent. Separate target bands are used for the 400 and 550 band levels (see **Supplementary Figure S2** online). Three or more out of four target bands was used as a threshold for the 400 and 550 band levels.

Ford method

Readers were asked to indicate whether target bands were present or absent. Separate target bands are used for the 400 and 550 band levels (see **Supplementary Figure S3** online). For both the 400 and 550 band levels, both target bands had to be present on both homologs to reach the threshold.

Comprehensive band counting method

Following the completion of data collection for the comparison of the five band-level estimation methods, four readers from four study centers (ARUP Cytogenetics Laboratory, Mayo Clinic Cytogenetics Laboratory, Cleveland Clinic Cytogenetics Laboratory, and University of Washington Cytogenetics Laboratory) completed an additional count of all bands on chromosomes 1–22 and chromosome X for each karyotype, omitting the acrocentric p arms. The lower count from two homologs was used as the haploid band length for each chromosome, and the sum of all bands on chromosomes 1–22 and X was calculated for the comprehensive haploid band level. This empirically determined band level was compared with the true band count shown in the 2013 International System for Human Cytogenetic Nomenclature idiograms for chromosomes 1–22 and X, omitting acrocentric p arms and counting the centromere of each chromosome as one band (349 bands for the 400 band level and 517 bands for the 550 band level).[6] Amniotic fluid karyotypes with average scores above 349 were designated as above the 400 band level, and peripheral blood karyotypes with average scores above 517 were designated as above the 550 band level.

Statistical analysis

The correlation between bands counted and reader experience was calculated using the Pearson correlation coefficient (r). The overall concordance between band-level estimation methods and between study centers was calculated using Cronbach's alpha statistical method and one-way ANOVA. All calculations were performed using Stata Statistical Software Release 12 (Stata, College Station, TX).

RESULTS

A total of 82 readers from 6 clinical cytogenetics laboratories participated in the study, together generating 2,000 scores for amniotic fluid karyotypes and 2,000 scores for peripheral blood karyotypes (**Table 1**). Readers spent an average of 76 min completing the entire survey. Readers' self-reported experience in human chromosome analysis ranged from 1 month to nearly 40 years, with a median of 8 years of experience. The total bands counted by a single reader for the entire study showed no correlation with that reader's experience (Pearson's $r = 0.03$; $P = 0.83$). Total bands counted was expressed as a ratio (total/total expected) equal to the sum total count for a single reader for the entire study, divided by the target sum total band count for the entire study (**Figure 1**). For readers who completed only half of the survey, the denominator (target sum total band count) was reduced accordingly. A reader who scored each karyotype by each method at exactly the cutoff threshold would thus have a total/total expected ratio of 1.

Five different band-level methods were assessed in this study by comparing the proportion of scores above the cutoff threshold for the desired band level. For amniotic fluid karyotypes, this threshold is 400 bphs, and for peripheral blood karyotypes, the threshold is 550 bphs. For each region evaluated, the lower score from the two homologs was used to generate a conservative value. For the five amniotic fluid karyotypes and the five peripheral blood karyotypes included in this study, the Welborn method was the most stringent, with the lowest fraction of scores reaching the cutoff threshold; the UKEQAS method was the least stringent method (**Table 2**). The overall concordance between methods was 94% as calculated by the Cronbach's alpha statistical method, and after eliminating the

Table 1 Survey participants

Study center	Number of survey participants	Median experience (range), in months
1	21	133 (29–432)
2	17	76 (13–240)[a]
3	14	25 (1–288)
4	13	132 (3–390)[a]
5	12	168 (6–324)[a]
6	3	364 (62–474)
7	2	(96)[a]
Total	82	96 (1–474)

[a]Reader experience not indicated for one or more of the participants. Of the 82 participants, 75 indicated their experience in clinical cytogenetics.

least concordant method (Welborn), the concordance between the remaining four methods increased to 97%. Using general linear mixed-effects models, there was a statistically significant difference between the methods ($P = 0.01$) (**Figure 2**). Analysis by sample type further showed that the difference between methods was statistically significant for the amniotic fluid karyotypes ($P < 0.01$) but not for the peripheral blood karyotypes ($P = 0.75$) (see **Supplementary Figures S4 and S5** online). Pairwise method comparisons, using the Bonferonni adjustment for multiple comparisons, indicated that the statistically significant difference was attributable to the Welborn method. Eliminating the Welborn method, the comparison of the remaining four methods showed no statistically significant

Figure 1 Band count ratio as a function of reader experience. The ratio displayed on the *y*-axis represents the sum total of the bands counted by one reader for the entire study (the total number of bands for all karyotypes and all methods) divided by the sum total expected for the designated target level (400 or 550 band level) for all karyotypes and all methods evaluated by that reader.

difference for either amniotic fluid karyotypes ($P = 0.62$) or peripheral blood karyotypes ($P = 0.72$).

The Stallard, Vancouver, and Ford methods were intermediate and fairly comparable. The UKEQAS method was the least stringent, yielding the most scores above the cutoff band level, and the Welborn method was the most stringent. As noted above, only the Welborn method was significantly different. Even though the methods differed somewhat in classifying the band level, the overall ordering of karyotypes by band level was fairly consistent from method to method. In other words, readers tended to agree on the hierarchy of the karyotypes from lowest to highest band level, and this agreement generally held up for each center and for each method (see **Supplementary Tables S1–S10** online).

A comprehensive count of all chromosome bands was also completed by four readers and was converted into a total band count designation (see **Supplementary Table S11** online). By this comprehensive method, two of the five amniotic fluid karyotypes (AF5 and AF9) and two of the five peripheral blood karyotypes (PB11 and PB15) were below the target band levels (400 bphs for amniotic fluid and 550 bphs for peripheral blood). The performance of the five different band-level estimation methods (Stallard, Vancouver, Welborn, UKEQAS, and Ford) relative to the comprehensive band count is shown in **Table 2**.

A statistical analysis was performed to analyze the differences between study centers with 10 or more participants. Scores generated by the different study centers were comparable for all centers except for laboratory 5, which generated consistently lower scores (see **Supplementary Tables S1–S10** online and **Figure 3**). There was a statistically significant difference between the distribution of scores obtained from different study centers as measured by a one-way ANOVA ($P < 0.01$). Using the Scheffe adjustment for multiple comparisons,
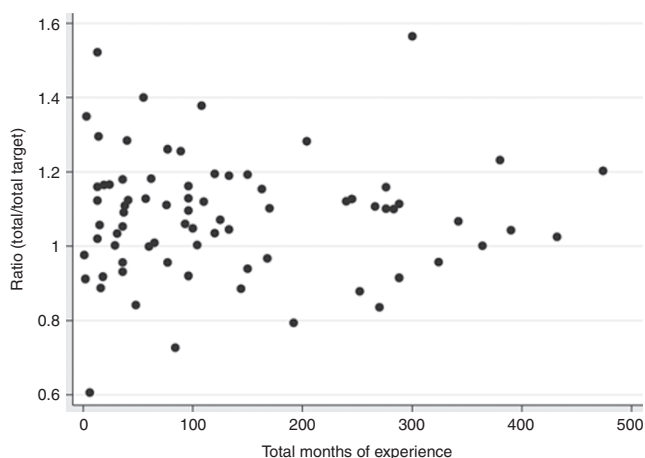
Table 2 Comparison of five band-resolution assessment methods

Karyotype ID	Stallard	Vancouver	Welborn	UKEQAS	Ford	Total band count designation
AF1	0.91	0.94	0.15	0.90	0.94	>400
AF5	0.46	0.33	0.04	0.74	0.33	<400
AF7	0.90	0.94	0.30	1.00	1.00	>400
AF8	0.84	0.93	0.14	0.99	0.96	>400
AF9	0.48	0.51	0.04	0.85	0.19	<400
PB4	0.58	0.63	0.65	0.85	0.85	>550
PB10	0.83	0.75	0.68	0.99	0.91	>550
PB11	0.08	0.15	0.10	0.30	0.10	<550
PB14	0.93	0.92	0.89	1.00	1.00	>550
PB15	0.35	0.23	0.16	0.68	0.34	<550
All	0.64	0.63	0.32	0.83	0.66	

The proportion of scores for each karyotype at or above the designated band-resolution threshold (400 bands per haploid set for amniotic fluid; 550 bands per haploid set for peripheral blood) is indicated. AF1, AF5, AF7, AF8, and AF9 are karyotypes from amniotic fluid specimens; PB4, PB10, PB11, PB14, and PB15 are karyotypes from peripheral blood specimens. The total band count designation in the last column is derived from the average total band count obtained from four readers from four different study centers (see **Supplementary Table S11** online). The average proportion of scores for all karyotypes is indicated in the last row of the table.

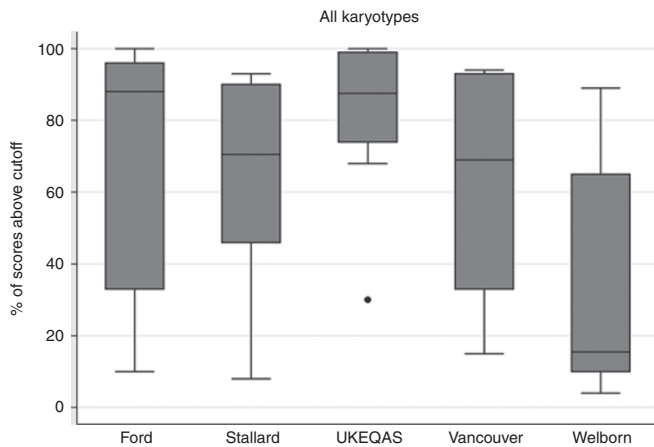UKEQAS, United Kingdom External Quality Assurance Scheme.

Figure 2 Box-and-whisker plot showing data distribution and interquartile ranges for each method with the median value displayed as a line in each box plot. Data from all scores of all karyotypes (amniotic fluid and peripheral blood) are represented here. An outlier in the UKEQAS score group (shown as a dot) had no impact on the results. UKEQAS, United Kingdom External Quality Assurance Scheme.
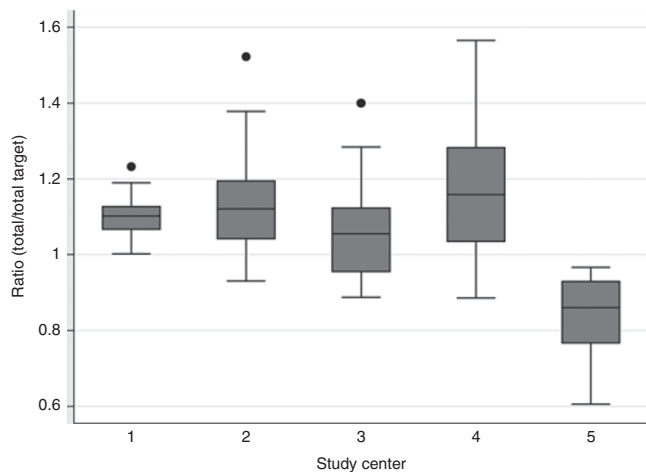

Figure 3 Box-and-whisker plot showing data distribution and interquartile ranges for each study center with the median value displayed as a line in each box plot. Outliers are shown as circles in the figure, but outliers had no impact on the results.

and eliminating laboratory 5 from the one-way ANOVA model, there was no statistically significant difference between the remaining centers ($P = 0.45$).

DISCUSSION

Five band-level assessment methods were compared for this study: three counting methods (Stallard, Vancouver, and Welborn) and two target band methods (UKEQAS and Ford). Each of the 5 methods was applied to the same 10 karyotypes (5 from amniotic fluid and 5 from peripheral blood) to determine if the karyotype reached the desired band level of 400 bphs for amniotic fluid or 550 bphs for peripheral blood.

Karyotypes for this study were chosen from digitally archived clinical cases of peripheral blood and amniotic fluid referred for the evaluation of the constitutional chromosome complement.

The primary criterion for karyotype selection was the presence of each target region uncrossed and free of debris or other artifact. This was more easily achieved for amniotic fluid karyotypes than for peripheral blood karyotypes, due to the greater length and crossing over of chromosomes in metaphase spreads from peripheral blood karyotypes. Therefore, the karyotypes were biased toward a lower band level. Because the target band level for amniotic fluid (400 bphs) and peripheral blood (550 bphs) is aimed at what is desirable but not always achievable for the specimen type, most of the karyotypes were in the borderline range for the desired band level. Selection of higher quality (i.e., higher band level) karyotypes would have yielded a better interobserver agreement in this study because a higher proportion of scores would have been above the cutoff threshold. However, for the sake of method comparison, average-to-borderline band level is better for maximizing differences between the methods.

A limitation of this analysis is that the readers received no prior instruction in the scoring methods, other than the instructions provided with the survey. It is therefore possible that lack of experience with a particular method may have skewed the results such that more complex methods were not evaluated as accurately. A further limitation of the study is that the survey participants were entrusted to complete the survey accurately, and occasional mistakes were likely present in the study data due to misunderstanding of the instructions. Every effort was made to simplify the survey form to ensure that data entry by primary participants was as accurate as possible. Data entry was manual (hand written) for the participants, and this was transcribed into an Excel database and independently verified by a second individual; however, data-entry errors are also possible with this study design. Another limitation of this study was that there is no true gold standard by which to compare the different band-level methods for accuracy. The true band level of each karyotype was estimated empirically by counting all of the bands on all of the chromosomes to generate a total band count. However, this comprehensive band-counting method was not entirely accurate because in almost every karyotype, regions of one or more chromosomes overlapped with other chromosomes. Moreover, this total band count is subjective like any other method, as demonstrated by the results obtained from the four different readers who completed this section. The subjectivity of band counting is evident in each section of this study given the range of the scores obtained from different readers. This variability in band-level assessment has no apparent dependence on experience in chromosome evaluation. By contrast, the variability appears to relate to the inherent stringency or permissiveness of the individual in terms of documenting the presence or absence of a visually subjective phenomenon. The finding of a significantly lower clustering of scores from laboratory 5 may indicate an overall trend toward conservatism in that laboratory; however, the remaining centers showed no significant difference in score clustering. Despite the subjectivity of chromosome band-level assessment, it is notable that readers generally agreed on the

ORIGINAL RESEARCH ARTICLE

hierarchy of karyotypes from lowest to highest band level. Stated in another way, the absolute band level assigned by each reader varied more than the overall ranking of karyotypes from lowest to highest band level. Therefore, it appears from this study that the choice of band-level assessment method is less critical than the consistency of use of a particular method, to allow laboratories to track the overall quality of karyotypes over time.

Chromosome analysis is based on visual inspection of chromosome bands, and although the International System for Human Cytogenetic Nomenclature band nomenclature provides a stable framework on which to evaluate an individual karyotype, we have shown that the band-level estimation methods are all inherently subjective. It is therefore perhaps more difficult than many cytogeneticists realize to define objective quality parameters for such purposes as clinical laboratory proficiency testing. The primary quality indicator of a karyotype is the band level, and as indicated by this study, the band level is subjective. The results obtained in this study suggest that readers agree better on relative band level than on absolute band level. This study also shows that the band-level estimation methods do not perform equally in discriminating lower from higher band level. Nevertheless, the lack of easy quality indicators does not negate the value of the karyotype for whole-genome analysis. Chromosome analysis continues to be a useful method for evaluating the large-scale architecture of the genome. Despite the revolution of new data obtainable by next-generation sequencing and chromosomal microarray, G-banded chromosome analysis endures as a vital laboratory technique.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

REFERENCES

1. Lawce H, Brown M. Cytogenetics: an overview. In: Barch M, Knutsen T, Spurbeck J (eds). *The AGT Cytogenetics Laboratory Manual, 2nd edn*. Lippincott-Raven Publishers; Philadelphia, 1997:19–75.
2. American College of Medical Genetics and Genomics. *Standards and Guidelines for Clinical Genetics Laboratories. E: Clinical Cytogenetics*. American College of Medical Genetics and Genomics: Bethesda, MD, 2010.
3. College of American Pathologists. *CAP Laboratory Accreditation Program. Cytogenetics Checklist*. College of American Pathologists: Northfield, IL, 2010.
4. Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 2010;86:749–764.
5. Manning M, Hudgins L; Professional Practice and Guidelines Committee. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet Med* 2010;12:742–745.
6. Shaffer LG, McGowan-Jordan J, Schmid M (eds). *ISCN (2013): An International System for Human Cytogenetic Nomenclature*. Karger: Basel, Switzerland,2013:16–31.
7. Human (*Homo sapiens*) Genome Browser Gateway. The Regents of the University of California. http://genome.ucsc.edu/cgi-bin/hgGateway. Accessed 15 May 2013.
8. Stallard R, Johnson W. Nonsubjective method for estimating the resolution of banded chromosomes. *Am J Hum Genet* 1983;35:155A.
9. Josifek K, Haessig C, Pantzar T. Evaluation of chromosome banding resolution: a simple guide for laboratory quality assurance. *Applied Cytogenet* 1991;17:101–105.
10. Welborn JL, Welborn R. Banding resolution of human chromosomes: a method of accuracy and simplicity. *Am J Med Genet* 1993;47:1180–1183.
11. Association for Clinical Cytogenetics (ACC) Professional Guidelines for Clinical Cytogenetics. *General Best Practice Guidelines (2007) v1.04*. British Society for Human Genetics: Birmingham, England, 2007.
12. Zabawski J, Wiktor A, Sikora M, van Dyke DL. Use reference bands to accurately estimate ISCN band levels 400, 500, and 850. *J Assoc Genet Technol* 2005;31:9–13.